6-2013

Demo: Approximate Semantic Matching in the COLLIDER Event Processing Engine

Souleiman Hasan

Kalpa Gunaratna
Wright State University - Main Campus, gunaratna.2@wright.edu

Yongrui Qin

Edward Curry

Repository Citation

Hasan, S., Gunaratna, K., Qin, Y., & Curry, E. (2013). Demo: Approximate Semantic Matching in the COLLIDER Event Processing Engine. *Proceedings of the 7th ACM International Conference on Distributed Event-Based Systems*, 337-338.
https://corescholar.libraries.wright.edu/knoesis/1036
10.1145/2488222.2489277

Demo: Approximate Semantic Matching in the COLLIDER Event Processing Engine

Souleiman Hasan
Digital Enterprise Research Institute
National University of Ireland, Galway
souleiman.hasan@deri.org

Kalpa Gunaratna
Kno.e.sis Center
Wright State University
Dayton OH
USA
kalpa@knoesis.org

Yongrui Qin
School of Computer Science
University of Adelaide, SA
Australia
yongrui.qin@adelaide.edu.au

Edward Curry
Digital Enterprise Research Institute
National University of Ireland, Galway
ed.curry@deri.org

ABSTRACT

This demo presents a use case from the energy management domain. It builds upon previous work on approximate semantic matching of heterogeneous events and compares two semantic matching scenarios: exact and approximate. It illustrates how a large number of exact matching event subscriptions are needed to match heterogeneous power consumption events. It then demonstrates how a small number of approximate semantic matching subscriptions are needed but possibly with a lower true positives/negatives performance. The demo is delivered via the COLLIDER approximate event processing engine currently under development in DERI.

Categories and Subject Descriptors

D.2.12 [**Software Engineering**]: Interoperability---*data mapping, interface definition languages*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval---*information filtering*.

Keywords

Approximate Event Matching; Loose Semantic Coupling; Semantic Matching.

1. INTRODUCTION

Event-based systems are decoupled in space, time, and synchronization which supports scalability [2]. However, scaling out to include participants from diverse domains poses a significant challenge in terms of the semantic interpretation of events. Current systems assume the existence of a mutual agreement between event producers and consumers on event semantics, which adds explicit dependencies between participants. In large-scale scenarios, such as the Internet-of-Things, high levels of semantic heterogeneity exist among events where different terms may be used to describe same concepts or event types.

If event consumers, or subscribers, are coupled to the underlying schema of events, they need to write many exact matching subscriptions to address the semantic heterogeneity in the events. We have previously proposed [3] an approximate semantic matching paradigm for event processing systems to work in such scenarios. This approach is implemented in the COLLIDER event processing engine which focuses on semantic approximation and its implications on event processing systems, especially on event enrichment and complex event processing.

Using approximate semantic event processing, a relatively small number of subscriptions are needed to match heterogeneous events. This facilitates the scalability of the system and the adoption by non-technical users. However, approximate matching may result in some false positives/negatives due to approximation. This demo focuses on these aspects of approximate semantic event matching within an energy management scenario.

2. DEMONSTRATION

The demo targets energy management within an office scenario where a user is interested in matching power consumption events in order to take energy saving actions [1]. Heterogeneity exists due to different sensors' manufacturers each using their own event description. The demo shows how a single subscription is used to match several event types using approximate matching semantics, whereas exact matching requires several subscriptions, typically one for each event type to achieve the same result.

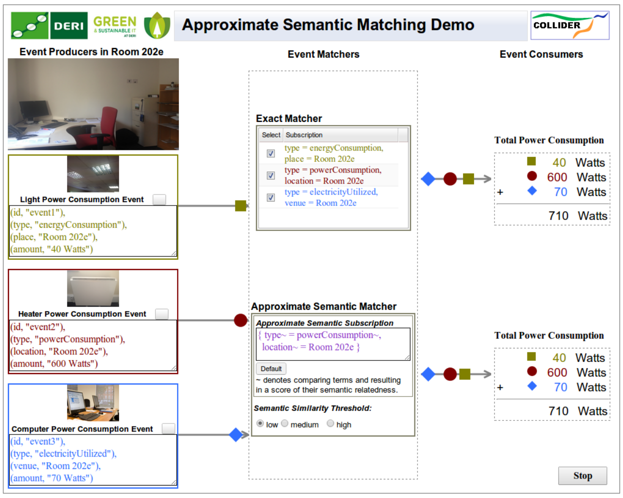

Figure 1. A screenshot from the demo application

2.1 Demo Setup

The main components of the demo are the set of events, the subscriptions and the matchers.

2.1.1 Events
There are three types of power consumption events, namely from sensors monitoring a light, a heater and a laptop. Each event has slightly different properties and descriptions. For example, the event types of *event 1, event 2, and event 3* are described using the terms *"energyConsumption"*, *"powerConsumption"*, and *"electricityUtilized"* respectively. However, these event types are semantically similar to each other in the energy domain. An example is an energy consumption event of a light that exists in *"Room 202e"* and consumes *40 watts* as illustrated in Figure 1. It is represented as a set of attribute-value pairs as the following:

{(id, "event 1"), (type, "energyConsumption"), (place, "Room 202e"), (amount, "40 Watts")}.

2.1.2 Subscriptions
There are two types of subscriptions in the demo: exact matching subscriptions and approximate matching subscriptions. There are three exact subscriptions, one for each event, and one approximate subscription for all the events. The exact matching subscriptions that target *event 1*, *event 2*, and *event 3* are respectively:

{type=energyConsumption, place=Room 202e}.

{type=powerConsumption, location=Room 202e}.

{type= electricityUtilized, venue=Room 202e}.

The approximate subscription is expressed in a COLLIDER language variant for attribute-value pairs as shown in Figure 1. It contains the *tilde* operator ~ after attributes and values to dictate that the approximate matcher shall match similar terms to the specified terms in the subscription. The approximate matching subscription targets all events and uses *powerConsumption~* to semantically match *"energyConsumption"*, *"powerConsumption"*, and *"electricityUtilized."* It also uses *location~* to semantically match the terms *place*, *location*, and *venue*. It is expressed as follows:

{type~ =powerConsumption~, location~ =Room 202e}.

2.1.3 Matcher Settings
There are two matchers behind the demo: an exact matcher and an approximate matcher as illustrated in Figure 1. The exact matcher can be set to match events against registered exact subscriptions. In order to match all three events, all three exact matching subscriptions must be registered in the exact matcher. The approximate matcher can be set with a matching threshold to determine how the events match the approximate subscription.

2.2 Demo Workflow
The objective of the use case is to identify the total power consumption of the events that match the user subscription(s). The basic steps in this demonstration application are as follows.

1. The user selects the exact matching subscriptions and defines an approximate matching subscription for the matchers.
2. The user selects and forward plays events to the exact and approximate matchers.
3. The exact matcher uses a subscription to match each event making the total power consumption matched 710 watts.
4. The approximate matcher obtains the correct total amount of power consumption of 710 watts when the threshold is low. If the threshold is medium or high the matched events become less and the total power consumption of matched events becomes less than 710 watts.

3. COLLIDER ARCHITECTURE
The approximation functionality of the demo is delivered via the COLLIDER event processing engine. COLLIDER is currently developed in DERI to study the various aspects of approximation in event processing and the potential impacts of uncertain matching on event enrichment and complex event processing.

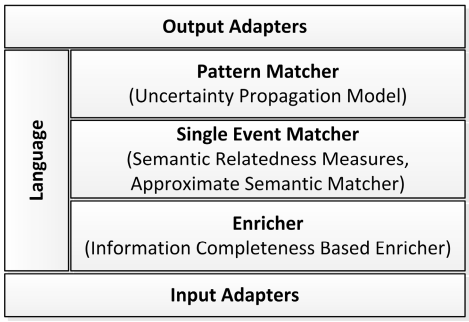

Figure 2. The COLLIDER engine architecture

Figure 2 shows a high level architecture of COLLIDER. Its components are briefly discussed in the following:

1. **Input and Output Adapters:** They allow the engine to interact with event producers and consumers respectively.
2. **The Language Module:** It is responsible for parsing the users' event subscriptions and complex event patterns.
3. **Enricher**: It enriches events which arrive at the engine to improve their information completeness with respect to potential event subscriptions.
4. **Single Event Matcher:** Approximate matching is enabled via a set of semantic relatedness measures as described in [3]. These measures are used to score the similarity between an event and a subscription based on the scores of each pair of terms. Approximate matching results in uncertain values that propagate to the pattern matcher.
5. **Pattern Matcher:** It propagates uncertainties from single event matching to obtain the uncertainty values for derived complex events.

This demo leverages the capability of the approximate single event matcher in COLLIDER while the other modules are out of the scope of this paper.

4. ACKNOWLEDGMENTS
This work has been funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

5. REFERENCES
1. Curry, E., Hasan, S., and O'Riain, S. Enterprise Energy Management using a Linked Dataspace for Energy Intelligence. *Second IFIP Conference on Sustainable Internet and ICT for Sustainability*, IEEE (2012).
2. Eugster, P.T., Felber, P.A., Guerraoui, R., and Kermarrec, A.M. The many faces of publish/subscribe. *ACM Computing Surveys (CSUR) 35*, 2 (2003), 114–131.
3. Hasan, S., O'Riain, S., and Curry, E. Approximate Semantic Matching of Heterogeneous Events. *6th ACM International Conference on Distributed Event-Based Systems (DEBS 2012)*, ACM (2012), 252–263.